



INTERPRETING BOX PLOTS – DATA ON CAMPING AND BACKPACKING GOODS

TEACHER VERSION

Subject Level:

Middle School Math

Grade Level:

6

Approx. Time Required:

60 minutes

Learning Objectives:

- Students will be able to create and interpret a box plot of census data.
- Students will be able to display numerical data on a number line.
- Students will be able to identify measures of center (median) and variability (interquartile range).

Activity Description

Students will review Economic Census data on the number of discount department stores in the United States that sold camping and backpacking equipment in 2007 so they can find the 5-number summaries and create box plots, using techniques such as forming a human number line and folding data sets on strips of paper.

Suggested Grade Level:

6

Approximate Time Required:

60 minutes

Learning Objectives:

- Students will be able to create and interpret a box plot of census data.
 - Students will be able to display numerical data on a number line.
 - Students will be able to identify measures of center (median) and variability (interquartile range).
-

Topics:

- 5-number summary
- Box plots
- Interquartile range
- Median
- Outliers
- Quartiles

Skills Taught:

- Analyzing data
 - Calculating a 5-number summary
 - Creating a box plot
-

Materials Required

- The student version of this activity, 5 pages
- A printed version of the data cards in the teacher version of the activity
- Masking tape
- Scissors for at least each pair of students

Activity Item

The following item is part of this activity. The item, its data source, and instructions for viewing the source data online appear at the end of this teacher version.

- Item 1: Group A and Group B Data Strips

For more information to help you introduce your students to the U.S. Census Bureau, read [*"Census Bureau 101 for Students."*](#) This information sheet can be printed and passed out to your students as well.

Standards Addressed

See charts below. For more information, read

[*"Overview of Education Standards and Guidelines Addressed in Statistics in Schools Activities."*](#)

Common Core State Standards for Mathematics

Standard	Domain	Cluster
CCSS.MATH.CONTENT.6.SP.B.4 Display numerical data in plots on a number line, including dot plots, histograms, and box plots.	6 SP – Statistics & Probability	Summarize and describe distributions.
CCSS.MATH.CONTENT.6.SP.B.5 Summarize numerical data sets in relation to their context, such as by:	6 SP – Statistics & Probability	Summarize and describe distributions.
CCSS.MATH.CONTENT.6.SP.B.5.C Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.		

Common Core State Standards for Mathematical Practice

Standard

CCSS.MATH.PRACTICE.MP4. Model with mathematics.

Students will create a box plot to represent a distribution of data on camping and backpacking equipment.

CCSS.MATH.PRACTICE.MP5. Use appropriate tools strategically.

Students will use number lines and fold data sets on strips of paper to create box plots.

National Council of Teachers of Mathematics' Principles and Standards for School Mathematics

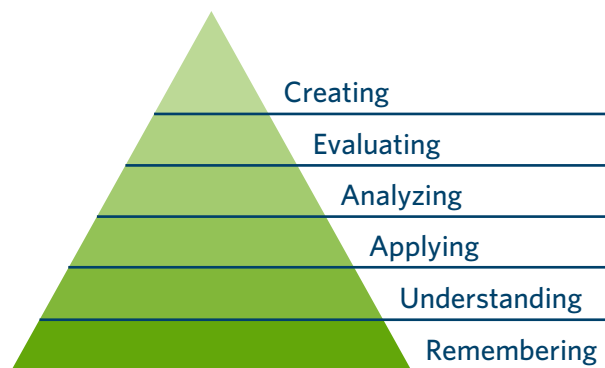
Content Standard	Students should be able to:	Expectation for Grade Band
Data Analysis and Probability	Formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them.	Select, create, and use appropriate graphical representations of data, including histograms, box plots, and scatterplots.
Data Analysis and Probability	Select and use appropriate statistical methods to analyze data.	Discuss and understand the correspondence between data sets and their graphical representations, especially histograms, stem-and-leaf plots, box plots, and scatterplots.

Guidelines for Assessment and Instruction in Statistics Education

GAISE	Level A	Level B	Level C
Formulate Questions	X		
Collect Data			
Analyze Data		X	
Interpret Results	X		

Bloom's Taxonomy

Students will **understand** how to find the values in a 5-number summary and **apply** this understanding to **create** and **analyze** box plots.



Teacher Notes

Before the Activity

Students must understand the following key terms:

- **Distribution** – an arrangement of a particular variable’s values that shows the frequency of each value’s occurrence in a data set and that can be described by its center, shape, and spread
- **Mean** – a measure of center in a set of numerical data, computed by adding the values in a list and then dividing by the number of values in the list
- **Median** – a measure of center in a set of numerical data, identified as the value appearing at the middle of a sorted version of the list (or the mean of the two middle values if the sorted list contains an even number of values)
- **Mode** – the value that appears most frequently in a data set
- **First quartile (Q_1)** – also known as lower quartile, the value that divides a sorted data set into the smallest 25 percent of the data and the largest 75 percent
- **Third quartile (Q_3)** – also known as upper quartile, the value that divides a sorted data set into the smallest 75 percent of the data and the largest 25 percent
- **Interquartile range (IQR)** – a measure of variability in a set of numerical data that is the distance between the first and third quartiles of the data set and that encompasses half of the data points (i.e., the difference between Q_3 and Q_1 in a five-number summary)
- **5-number summary** – a summary that provides information about a data set and contains five values: the minimum (smallest value), first quartile, median (middle value), third quartile, and maximum (largest value)
- **Measure of center** – a value that describes the central focus of a data set, commonly represented as the mean or median
- **Measure of variability** – a measure of how spread out the values are in a data set
- **Range** – the numerical difference between a data set’s maximum value and minimum value

Students should have the following skill:

- Ability to calculate an average of two numbers

Teachers should create a straight line on the floor with masking tape to serve as a number line large enough for students to position themselves on. Its values should go from 0 to 100, with labels at every fifth interval. Teachers should also cut out the following data cards, placing them in a container from which students can draw them.

Data Cards: Number of Discount Department Stores in Selected U.S. States That Sold Camping and Backpacking Equipment: 2007

Alabama	46	Kentucky	54	Oklahoma	33
Alaska	6	Louisiana	26	Oregon	48
Arizona	69	Maine	21	Rhode Island	11
Arkansas	29	Maryland	83	South Carolina	54
Colorado	48	Massachusetts	87	South Dakota	19
Connecticut	51	Mississippi	19	Tennessee	60
Delaware	10	Missouri	85	Utah	39
Georgia	81	Montana	25	Vermont	6
Hawaii	15	Nebraska	28	Virginia	97
Idaho	25	Nevada	26	Washington	81
Indiana	80	New Hampshire	33	West Virginia	18
Iowa	52	New Mexico	27	Wyoming	11
Kansas	32	North Dakota	12		

Source: 2007 Economic Census

Teachers should introduce the activity by asking students what kind of equipment is needed for camping or backpacking, whether they have ever purchased such equipment, and, if so, where those purchases were made.

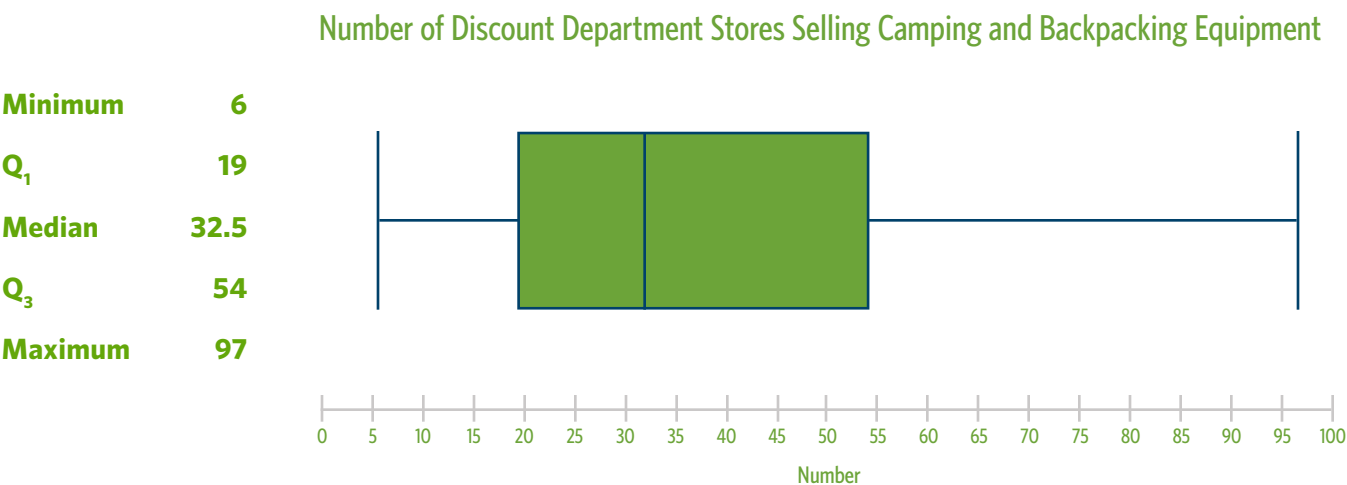
For a warmup, teachers should have students form a human number line to help them visualize a box plot. As each student draws a data card from the container, teachers should explain what the cards show and direct students to position themselves on the corresponding value on the number line. (Teachers should be aware that not all 38 data cards are needed for this activity.)

Recording the values on the board as they go, teachers should have students identify the people on their human number line who represent:

- The minimum and maximum
- The median, as *either*:
 - The value for the person in the middle (who can go back to his or her seat afterward to ensure exclusion from Q_1 and Q_3 calculations)
 - The average of the two values for the two students in the middle (who should stay on the number line because these values are included in Q_1 and Q_3)
- Q_1 and Q_3

During this exercise, teachers should help students understand that when a data set lacks a single data point for its median, the two middle data points belong to different halves of the data and must be used to calculate Q_1 and Q_3 .

Once students return to their seats, teachers should use the recorded data to draw a box plot on the board, emphasizing the values in the five-number summary. If all data cards are used, the summary and box plot should look like:



Teachers should then have students cut out both data strips in **Item 1**. To complete the activity, teachers should have students work either in pairs, with each partner using a single data strip, or individually, using both data strips.

During the Activity

Teachers should monitor students as they work.

After the Activity

Teachers should review students' answers to question 7 and lead a discussion about outliers and their effect on a data distribution, as shown with the data for California. Teachers could also ask students to compare and contrast the box plots they created with a partner's box plots and then share their observations with the class.

Extension Ideas

- Teachers could have students make box plots and/or 5-number summaries electronically using Microsoft Excel or an applet like Interactivate (www.shodor.org/interactivate/activities/BoxPlot).
- Teachers could have students compare and contrast two box plots of different data sets, such as sporting goods stores (e.g., Bass Pro Shops) and outdoor equipment stores (e.g., REI).

Student Activity

Click [here](#) to download a printable version for students.

Activity Item

The following item is part of this activity and appears at the end of this student version.

- Item 1: Group A and Group B Data Strips

Student Learning Objectives

- I will be able to create and interpret a box plot of census data.
- I will be able to display numerical data on a number line.
- I will be able to identify measures of center (median) and variability (interquartile range).

- Follow these steps to help you fill in the blanks for the 5-number summary and to create a box plot, adding a title and horizontal axis label, of the data in **Item 1: Group A and Group B Data Strips**:

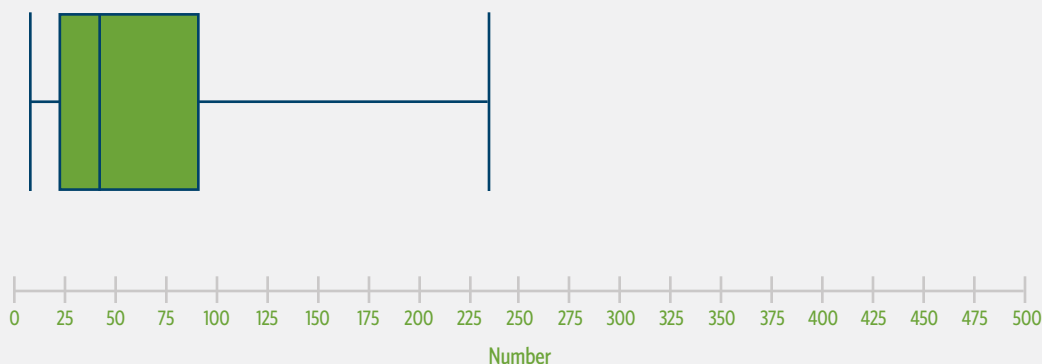
Fold a strip in half to find the median. Then fold it in half again to find the lower quartile (Q_1) and upper quartile (Q_3). If these folds are ever between two numbers, average those two numbers to calculate the value you need, rounding to one decimal place.

Group A (excludes California)

Minimum = 6 Q_1 = 25 Median = 48 Q_3 = 92 Maximum = 229

Student box plots should look like:

Number of Discount Department Stores Selling Camping and Backpacking Equipment in U.S. States (excluding California): 2007

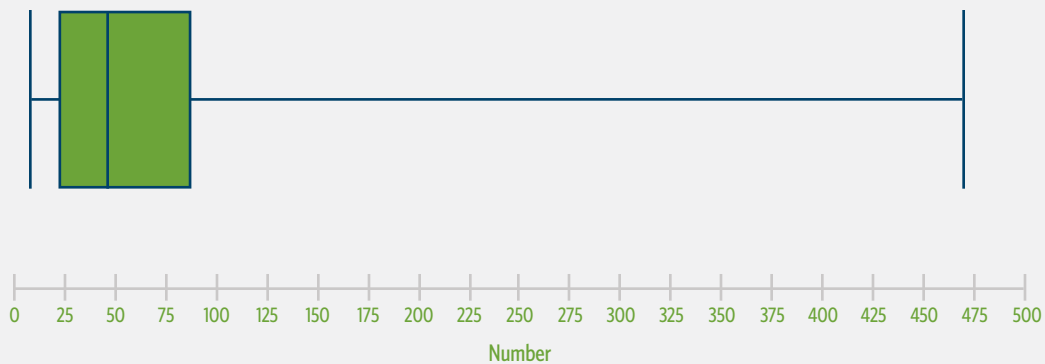


Group B (includes California)

Minimum = 6 Q_1 = 25 Median = 49.5 Q_3 = 97 Maximum = 470

Student box plots should look like:

Number of Discount Department Stores Selling Camping and Backpacking Equipment in U.S. States (including California): 2007



2. Why must the data be in numerical order before you start folding your paper to find the 5-number summary?

Because Q_1 , the median, and Q_3 divide an ordered data set into four equal parts (quarters). For example, the median separates the smallest half of the data from the largest half. If the data are not in order, folding the list in half won't necessarily tell you the value of the median.

3. Which quarters of the data sets have the most states?
 - Group A: **All quarters have the same number of states.**
 - Group B: **All quarters have the same number of states.**

4. Calculate the range for each quarter by filling in the following table (showing your work).

Quarter and Its Range Formula	Group A	Group B
First (lowest) = Q_1 - minimum	25 - 6 = 19 stores	25 - 6 = 19 stores
Second = median - Q_1	48 - 25 = 23 stores	49.5 - 25 = 24.5 stores
Third = Q_3 - median	92 - 48 = 44 stores	97 - 49.5 = 47.5 stores
Fourth (highest) = maximum - Q_3	229 - 92 = 137 stores	470 - 97 = 373 stores

5. The interquartile range (IQR) is the difference between Q_3 and Q_1 . Find the IQR for both data sets (showing your work).

- Group A: **92 - 25 = 67 stores**
- Group B: **97 - 25 = 72 stores**

6. Three things to examine in a data distribution are center, shape, and spread. Look at your box plot to describe those three characteristics for both sets of data in the following table.

Characteristic	Group A	Group B
Center – most of the data are centered at:	48 stores	49.5 stores
Shape – the quarter where the data values are more spread out is:	Fourth	Fourth
Spread (show your work) – the range of the data is:	229 - 6 = 223 stores	470 - 6 = 464 stores

7. The data point for California is considered an outlier in this data set. Why do you think that is?

Student answers will vary, but the correct answer is that the number of discount department stores in California that sold camping and backpacking supplies (470) is much higher than the value for any other state.

8. Describe the main differences between the box plots for Groups A and B.

Student answers will vary but could include: The maximum for Group B is much higher than the maximum for Group A, so the right whisker is much longer for that plot.

9. When looking at a data distribution, it is helpful to know what pieces of the data set you can see. Which of the following values and measures does a box plot show? (Check all that apply.)

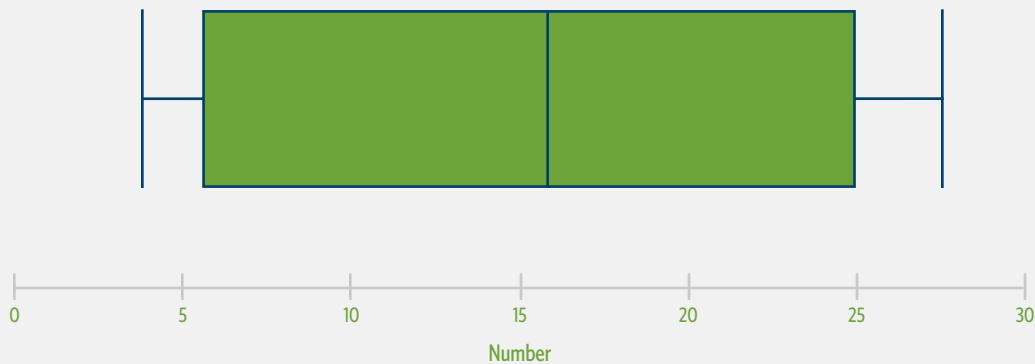
- ☒ the minimum
☐ the mean
☒ the maximum
☐ the seventh data point
☒ the median
☐ the mode

10. Now apply what you've learned to interpret the distribution of a new data set. Calculate the 5-number summary and make a box plot for these values: 4, 7, 16, 23, 27.

Minimum = 4 Q_1 = 5.5 Median = 16 Q_3 = 25 Maximum = 27



Student box plots should look like:



11. Use what you found in question 10 to interpret the distribution of the data set.

Student answers will vary but could include that one-fourth of all data points are between 4 and 5.5, between 5.5 and 16, between 16 and 25, and between 25 and 27.

Item 1: Group A and Group B Data Strips

Number of Discount Department Stores Selling Camping and Backpacking Equipment in 2007 for Each U.S. State
(from least to greatest)

Group A (excludes California)

Alaska	6
Vermont	6
Delaware	10
Rhode Island	11
Wyoming	11
North Dakota	12
Hawaii	15
West Virginia	18
Mississippi	19
South Dakota	19
Maine	21
Idaho	25
Montana	25
Louisiana	26
Nevada	26
New Mexico	27
Nebraska	28
Arkansas	29
Kansas	32
New Hampshire	33
Oklahoma	33
Utah	39
Alabama	46
Colorado	48
Oregon	48
Connecticut	51
Iowa	52
Kentucky	54
South Carolina	54
Tennessee	60
Arizona	69
Indiana	80
Georgia	81
Washington	81
Maryland	83
Missouri	85
Massachusetts	87
Virginia	97
North Carolina	107
Minnesota	114
New Jersey	117
Wisconsin	133
New York	156
Michigan	158
Ohio	158
Texas	163
Pennsylvania	176
Illinois	207
Florida	229

Group B (includes California)

Alaska	6
Vermont	6
Delaware	10
Rhode Island	11
Wyoming	11
North Dakota	12
Hawaii	15
West Virginia	18
Mississippi	19
South Dakota	19
Maine	21
Idaho	25
Montana	25
Louisiana	26
Nevada	26
New Mexico	27
Nebraska	28
Arkansas	29
Kansas	32
New Hampshire	33
Oklahoma	33
Utah	39
Alabama	46
Colorado	48
Oregon	48
Connecticut	51
Iowa	52
Kentucky	54
South Carolina	54
Tennessee	60
Arizona	69
Indiana	80
Georgia	81
Washington	81
Maryland	83
Missouri	85
Massachusetts	87
Virginia	97
North Carolina	107
Minnesota	114
New Jersey	117
Wisconsin	133
New York	156
Michigan	158
Ohio	158
Texas	163
Pennsylvania	176
Illinois	207
Florida	229
California	470

Source: 2007 Economic Census